

ABSTRACT OF THE DISCLOSURE

[00176] A method of searching a document having nested-structure document-specific markup (such as Extensible Markup Language (XML)) involves 112 receiving a query that designates at least (A) a phrase to be matched in a phrase matching process, and (B) a selective designation of at least a tag or annotation that is to be ignored during the phrase matching process. The method further involves 114 deriving query-specific indices based on query-independent indices that were created specific to each document, and 116 carrying out the phrase matching process using the query-specific indices on the document having the nested-structure document-specific markup. The query-independent indices may be created by a method including 104 labeling elements in the document with intervals, in which (1) for markup tags, the intervals are defined in terms of a starting index number associated with an opening markup tag and an ending index number associated with a closing markup tag that corresponds to the opening markup tag, and in which (2) for single words, the intervals are defined in terms of a single index number associated with the word. The method further involves 106 forming the query-independent indices 206 so that they are configured to be used in the searching method 116 by first receiving, for a word or tag in the document, a position in the document, and by then indicating whether or not the word or tag is present at that position.